A quick tutorial on PDF Transformation

Written by: Chieh Wu

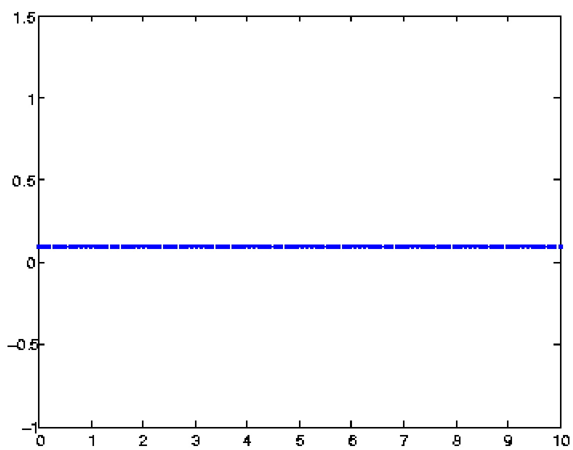
This tutorial assumes:

- The reader already understand the concept of PDF, CDF, random variables.

This tutorial will:

- Show you how to transform a PDF function into another one
- How to generate random variables using the uniform distribution.

Let us start with a random variable X. This random variable can be anything. For the sake of our example let us use a uniform distribution. Let's say we are waiting for the arrival of a bus. The bus can show up between 6:00 and 6:10. The distribution for the bus to show up is therefore uniform between 10 minutes with the probability of 1/10.



Now let's assume we have a PDF of Y such that Y = 2X. To give you a more concrete idea of what the equation Y=2X means, it is now basically saying that the bus can show up anywhere between 6:00 to 6:20. Before the possible outcomes of X were between 0 to 10 minutes. Now the possible outcomes of Y is two times the outcome of X. We therefore have 20 minutes. The goal of this paper is to show how we can find the PDF of Y by knowing X and the relationship between X and Y.

Of course, this is an oversimplified example since most probability students can guess that a uniform distribution would now have a probability of $\frac{1}{20\,\text{minutes}}$. By going through this exercise, the goal is to understand the mechanism of how we can transform any continuous distribution.

Let us once again start with a distribution of X. The distribution of X is related to a distribution of Y such that $Y = f(X)$. The distribution of Y is also related to X such that $X = f^{-1}(Y)$.

In our case

$$y = f(x) \quad = 2x \qquad \text{equation: } 1$$
$$x = f^{-1}(y) = \tfrac{1}{2}y \qquad \text{equation: } 2$$

Before I go on further, let me explain the notation of X and x. The capital X is the random variable while x is the actual number within the random variable. So in our case, X is the random variable from 0 to 10 while x may be the number 5 within 10 minutes. The same idea goes with Y and y. The capital Y is the distribution while y is the actual variable.

The cdf of X and Y are therefore:

$$\mathrm{F}_x(x) = p(X \leq x) \qquad \text{equation: } 3$$
$$\mathrm{F}_y(y) = p(Y \leq y) \qquad \text{equation: } 4$$

If we plug x from equation 2 into x in equation 3 we would get:

$$\mathrm{F}_x(x) = p(\ f^{-1}(y)\ \leq x) = \ p(\tfrac{1}{2}y \leq x) = p(y \leq 2x) = \mathrm{F}_y(2x)$$

This is a very important transition. The key we want to show here is that

$$\mathrm{F}_x(x) = \mathrm{F}_y(2x) \quad \Rightarrow \ \mathrm{F}_x(\tfrac{1}{2}y) = \mathrm{F}_y(y) \quad \Rightarrow \qquad \mathrm{F}_x(f^{-1}(y)) = \mathrm{F}_y(y)$$

In english, this basically means that the CDF of X is related to the CDF of Y. In our example we know that the probability of x is $\frac{1}{10\,\text{minutes}}$. To find the CDF we would take the integral of the pdf.

$$\mathrm{F}_x(x) = \int \ \frac{1}{10}dx = \frac{1}{10}x$$

Now if we want to get the CDF of y we use the relationship

$$\mathrm{F}_x(f^{-1}(y)) = \mathrm{F}_y(y) \quad \Rightarrow \quad \frac{1}{10}\,(\,\frac{1}{2}y) \ = \ \frac{1}{20}\,y$$

Now if we want to convert from CDF back to PDF we would take the derivative of the CDF with respect to y.

$$\mathrm{P}_y(y) = \frac{1}{20}$$

From this let's now go over the step by step instruction on transformation.

1. Take the integral of the PDF you are transforming.
2. Plug the inverse relationship between X and Y into the CDF of X
3. Take the derivative of the CDF of X

Part 2

How do you generate different random variables using a 0 to 1 uniform distribution

From the equation above

$$F_x(x) = F_y(f^{-1}(y))$$

Since we are using the uniform distribution

$$F_x(x) = x$$

This means that we really have

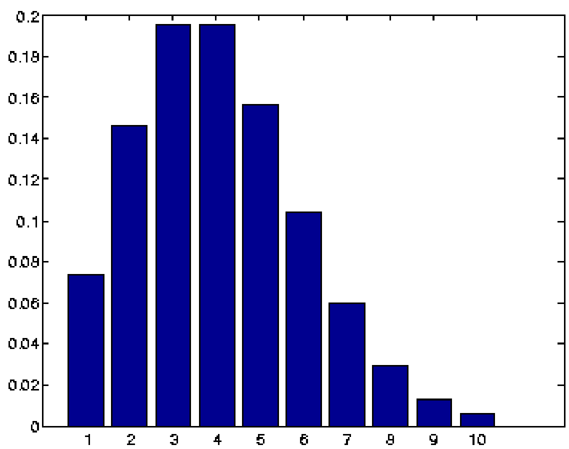
$$x = F_y(f^{-1}(y))$$

For example, let's say we want to generate a poisson random variable using a uniform distribution.

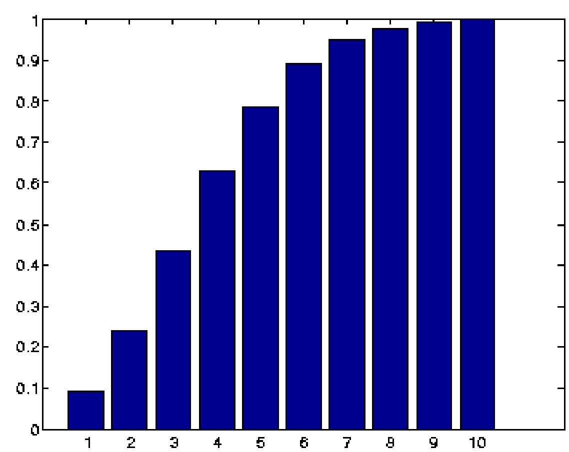
The formula for the poisson distribution is:

$$p(n) = e^{-\lambda}\frac{\lambda^n}{n!}$$

With λ of 4 we would have a distribution that looks like this



And a CDF that looks like this

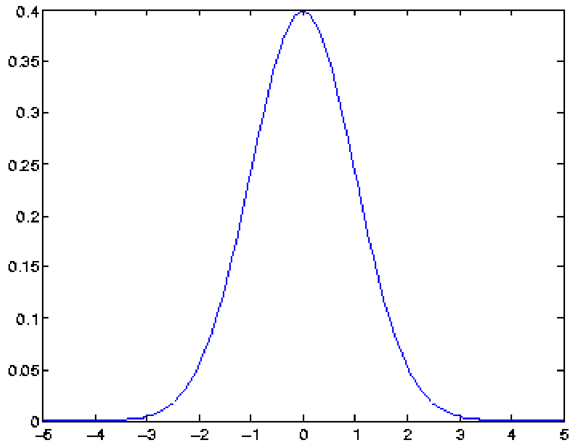


If you want to generate a poisson distribution, all you have to do is generate a random uniform variable from 0 to 1. This will represent the y axis on the CDF graph above. Once you have the y axis, you simply match it up with the closest corresponding x axis. So let's say if your uniform generator outputs 0.6, in the code you would use a look up table and find that you have really generated the number 4.

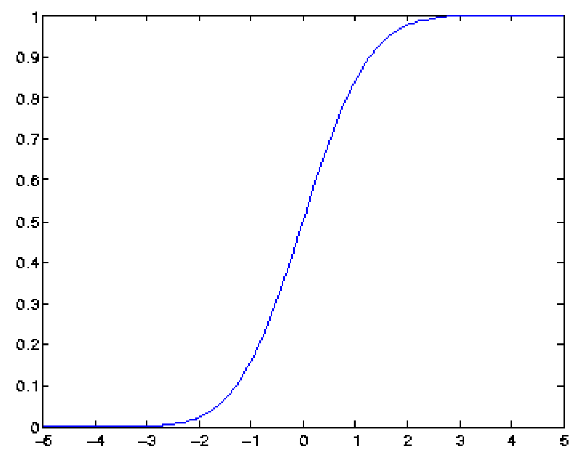
Let's take a look at a gaussian distribution as well.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}\frac{x^2}{\sigma^2}}$$

If we let the mean to be zero and variance equal to 1, we would have a graph like this.



The CDF of the gaussian



We would generate the gaussian distribution the same way. We first generate a random uniform number from zero to one. Then we'll find the the corresponding x axis on the CDF graph.

This is hopefully a help for some people. It is normally a pain to generate gaussian distribution in c or many other programming languages. I think I have messed up something in my derivation at the first section, I am probably not using the most perfect notation. if you noticed the mistake, please let me know.

I can be contacted at juliusctw@yahoo.com
You can also find my other writings and work at http://www.ourmedia.org/user/4887